A Novel Sparse Graphical Approach for Multimodal Brain Connectivity Inference

Bernard Ng, Gaël Varoquaux, Jean-Baptiste Poline, Bertrand Thirion

A Novel Sparse Graphical Approach for Multimodal Brain Connectivity Inference

Bernard Ng, Gael Varoquaux, Jean Baptiste Poline, and Bertrand Thirion

Parietal team, Neurospin, INRIA Saclay-Ile-de-France, France
bernardyng@gmail.com

Abstract. Despite the clear potential benefits of combining fMRI and diffusion MRI in learning the neural pathways that underlie brain functions, little methodological progress has been made in this direction. In this paper, we propose a novel multimodal integration approach based on sparse Gaussian graphical model for estimating brain connectivity. Casting functional connectivity estimation as a sparse inverse covariance learning problem, we adapt the level of sparse penalization on each connection based on its anatomical capacity for functional interactions. Functional connections with little anatomical support are thus more heavily penalized. For validation, we showed on real data collected from a cohort of 60 subjects that additionally modeling anatomical capacity significantly increases subject consistency in the detected connection patterns. Moreover, we demonstrated that incorporating a connectivity prior learned with our multimodal connectivity estimation approach improves activation detection.

Keywords: brain connectivity, diffusion MRI, fMRI, multimodal integration.

1 Introduction

Recent evidence suggests that the effects of many neurological diseases are manifested through abnormal changes in brain connectivity [1]. Techniques for inferring connectivity from functional magnetic resonance imaging (fMRI) data can be largely divided into two categories. One category comprises techniques, such as the seed-based approach and independent component analysis (ICA) [1], which group brain areas into networks. The other category includes techniques that estimate the connectivity between brain areas [2] and apply graph-theoretic measures to characterize the estimated connection structure. We focus on connectivity estimation in this work.

The strong noise inherent in fMRI data and its high dimensionality given the typically small sample sizes pose major challenges to reliable connectivity estimation [2]. Since neural dynamics is largely shaped by the structure of the underlying fiber pathways [3, 4], informing connectivity estimation with anatomical information extracted from e.g. diffusion MRI (dMRI) data should prove beneficial. Past studies that jointly examined dMRI and fMRI data have primarily focused on comparing connectivity measures estimated from each modality separately [3, 4]. The general finding is that brain areas with high anatomical connectivity typically exhibit high functional con-

nectivity [3, 4], but the converse does not necessarily hold due to factors, such as noise-induced correlations in fMRI observations, indirect functional connections, and tractography errors [4]. Other widely-used approaches for multimodal analysis include employing fMRI to guide seed selection in tractography [3, 4]. The use of fiber bundle shapes to predict activated brain areas has also been explored [5]. A major limitation to the aforementioned approaches is that pooling information extracted independently from each modality does not capitalize on the complementary information that a joint analysis of the two modalities would facilitate. To exploit the joint information in dMRI and fMRI data, variants of ICA and canonical correlation analysis [6] have been proposed for identifying brain areas that display high correlations between anatomical and functional attributes, such as fractional anisotropy and activation effects. Recently, a probabilistic model has been put forth for combining dMRI and fMRI data in detecting group differences in brain connection structure [7].

In this paper, we propose a novel multimodal integration approach based on sparse Gaussian graphical model (SGGM) for estimating intra-subject brain connectivity. Specifically, we cast connectivity estimation as a sparse inverse covariance learning problem [8]. Since elements of the inverse covariance matrix are proportional to the partial correlations between the associated variable pairs, zero entry would indicate conditional independence [8]. Using SGGM thus enables simultaneous estimation of connection strength and structure. To integrate anatomical information into functional connectivity estimation, we adapt the degree of sparse penalization on each functional link based on its anatomical capacity. Functional connections with less anatomical support are thus more heavily penalized, which helps reduce false detection of noise-induced functional connections. Also, using partial correlations as a measure of functional connectivity reduces the effects of indirect interactions. Furthermore, although larger penalizations are exerted on functional connections with no anatomical support, if ample evidence from fMRI data suggest the presence of such links, these connections will not necessarily be assigned zero connectivity. Our approach hence provides some tolerances to the inconsistencies between dMRI and fMRI-derived connectivity measures that hinder integration of these modalities. On a large dataset of 60 subjects, we showed that applying our multimodal approach significantly increases subject consistency in the detected connection structure over analyzing fMRI data alone. Enhanced group activation detection was also obtained by incorporating a connectivity prior [9] learned with our proposed approach, thus demonstrating the gain of integrating anatomical and functional information in brain connectivity estimation.

2 Methods

In this work, we focus on integrating dMRI with resting state (RS) fMRI for estimating brain connectivity. For this, we propose a SGGM-based approach (Section 2.1). Critical to the estimation is the choice on sparsity level, which we optimize using cross validation with a refined grid search strategy (Section 2.2). We validate our approach based on subject consistency and group activation detection (Section 2.3).

2.1 Sparse Gaussian Graphical Model

Let $\mathbf{S}$ be a $d \times d$ sample covariance matrix computed from data that are assumed to follow a centered multivariate Gaussian distribution. In the present context of brain connectivity estimation, $\mathbf{S}$ corresponds to correlations between the RS-fMRI observations of d brain areas of a given subject. To learn a well-conditioned sparse inverse covariance matrix, $\hat{\mathbf{\Lambda}}$, we minimize the negative log data likelihood over the space of positive definite (p.d.) matrices, $\mathbf{\Lambda} > 0$, with an l_1 penalty imposed on $\mathbf{\Lambda}$ [8]:

$$\hat{\mathbf{\Lambda}} = \arg\min_{\mathbf{\Lambda}>0} tr(\mathbf{S\Lambda}) - \log\det(\mathbf{\Lambda}) + \lambda \sum_{i=1}^{d}\sum_{j=1}^{d} \mathbf{W}_{ij}\left|\mathbf{\Lambda}_{ij}\right|, \quad (1)$$

where λ governs the overall level of sparsity and $\mathbf{W}_{ij}$ differentially weights the amount of sparse penalization on each connection based on its anatomical capacity:

$$\mathbf{W}_{ij} = \begin{cases} \exp\left(-\mathbf{K}_{ij}/\sigma\right) & i \neq j \\ 0 & i = j \end{cases}, \quad (2)$$

where $\mathbf{K}_{ij}$ corresponds to some measures of anatomical capacity. We set $\mathbf{K}_{ij}$ as the fiber count between brain areas i and j [3, 4] with the amount of penalization saturating to zero for $\mathbf{K}_{ij} >> \sigma$, modeling how the additional fibers may reflect redundant wiring. Other decreasing functions of $\mathbf{K}_{ij}$ can also serve as $\mathbf{W}_{ij}$. We defer the selection of $\mathbf{W}_{ij}$ for future work. σ is learned from data (Section 2.2). In accordance to past findings [3, 4], $\mathbf{\Lambda}_{ij}$ associated with brain areas that have fewer connecting fibers are more strongly penalized. Note that we have explicitly set $\mathbf{W}_{ij}$ to 0 for $i = j$, which has been theoretically proven and empirically shown to provide more accurate solutions of (1) [10]. To solve (1), we employ a recent second-order algorithm [8] that facilitates efficient computation of Newton steps with iterates guaranteed to remain p.d. This algorithm provides substantially faster convergence rate than current gradient-based methods. We define convergence as having a duality gap, η, below 10^{-5} [10]:

$$\eta = tr\left(\mathbf{S\Lambda}\right) + \lambda \sum_{i=1}^{d}\sum_{j=1}^{d} \mathbf{W}_{ij}\left|\mathbf{\Lambda}_{ij}\right| - d. \quad (3)$$

2.2 Parameter Selection

The estimated connection structure critically depends on λ. To learn the optimal λ and σ data-drivenly for each subject, we combine cross validation with a refined grid search strategy, as summarized in Algorithm 1. The algorithm requires the following

inputs: $\mathbf{Z}$ = an $t \times d$ matrix containing RS-fMRI time courses of d brain areas, $\mathbf{K}$ = fiber count matrix, λ_{lb} and λ_{ub} = initial lower and upper bounds of the λ search range, σ_{lb} and σ_{ub} = lower and upper bounds of the σ search range, R = number of refinement levels, F = number of grid points for λ, G = number of grid points for σ, and C = number of cross validation folds. We set λ_{ub} to $\max|\mathbf{S}_{ij}|$, $i \neq j$, which corresponds to the maximum λ beyond which $\hat{\mathbf{\Lambda}}_{ij}$ is guaranteed to shrink to 0 [10]. As for λ_{lb}, we empirically found that setting λ_{lb} to $\lambda_{ub}/100$ assigns non-zero value to >90% of the elements in $\hat{\mathbf{\Lambda}}$ in the case of uniform sparse penalization. We fix σ_{lb} and σ_{ub} to the 25th and 75th percentile of the fiber count distribution. R, F, G, and C are set to 3, 5, 5, and 3, respectively.

Algorithm 1 Refined Grid Search for λ_{opt} and σ_{opt}

Input: $\mathbf{Z}$, $\mathbf{K}$, λ_{lb}, λ_{ub}, σ_{lb}, σ_{ub}, R, F, G, C
Output: λ_{opt}, σ_{opt}
Temporally divide $\mathbf{Z}$ into C folds
Set σ_{grid} to a log grid with G grid points between σ_{lb} and σ_{ub}
for $r = 1$ to R
  Set λ_{grid} to a log grid with F grid points between λ_{lb} and λ_{ub} in decreasing order
  for $f = 1$ to F, $g = 1$ to G, and $c = 1$ to C
    Estimate sample covariance, $\mathbf{S}_{train}$, with all folds of $\mathbf{Z}$ except the c^{th} fold
    Estimate sample covariance, $\mathbf{S}_{test}$, with the c^{th} fold of $\mathbf{Z}$
    Solve (1) to find $\mathbf{\Lambda}_{train}$ with $\lambda = \lambda_{grid}(f)$, $\sigma = \sigma_{grid}(g)$
    Compute data likelihood, $dl(f,g,c)$, of $\mathbf{S}_{test}$ given $\mathbf{\Lambda}_{train}$, based on (1)
      without the sparse penalization term
  end
  Find λ_{opt} and σ_{opt} based on maximum of the average $dl(f,g,c)$ over C folds
  if $\lambda_{opt} == \lambda_{grid}(1)$
    Set λ_{lb} to $\lambda_{grid}(2)$, set λ_{ub} to $\lambda_{grid}(1)$
  else if $\lambda_{opt} == \lambda_{grid}(F)$
    Set λ_{lb} to $\lambda_{grid}(F)/10$, set λ_{ub} to $\lambda_{grid}(F\text{-}1)$
  else
    Find f_{opt} corresponding to λ_{opt}
    Set λ_{lb} to $\lambda_{grid}(f_{opt}\text{-}1)$, set λ_{lb} to $\lambda_{grid}(f_{opt}+1)$
  end
end

2.3 Validation

We base our validation on: 1) increased subject consistency in the support of $\hat{\mathbf{\Lambda}}$ and 2) increased group activation detection. The rationale behind the first criterion is that subjects within the same population should have similar brain connection structure [2]. As for the second criterion, we have shown in a previous work [9] that incorporating a RS-connectivity prior improves group activation detection. Thus, greater increase in group activation detection presumably implies that the corresponding connectivity estimates better reflect the underlying neural circuitry.

Subject Consistency. For each pair of subjects in our dataset (Section 3), we compare the support of their $\hat{\mathbf{\Lambda}}$'s using the Dice Similarity Coefficient (DSC), defined as TPR / (2TPR + FPR + FNR), with each subject alternately taken as the reference. TPR, FPR, and FNR are the true positive rate, false positive rate, and false negative rate, respectively. Both connections commonly present in each pair of subjects as well as those in one subject but not the other are thus accounted for in our chosen metric.

Group Activation Detection. Our previously proposed connectivity-informed activation model [9] is summarized below:

$$\mathbf{Y} \sim N(\mathbf{AX}, \mathbf{V}_1) = \frac{1}{\left|2\pi\mathbf{V}_1\right|^{\frac{n}{2}}} \exp\left(-\frac{1}{2} tr\left((\mathbf{Y}-\mathbf{AX})^T \mathbf{V}_1^{-1}(\mathbf{Y}-\mathbf{AX})\right)\right), \tag{4}$$

$$\mathbf{A} \sim MN(0, \mathbf{V}_2, \alpha\mathbf{X}\mathbf{X}^T) = \frac{\left|\alpha\mathbf{X}\mathbf{X}^T\right|^{\frac{d}{2}}}{\left|2\pi\mathbf{V}_2\right|^{\frac{m}{2}}} \exp\left(-\frac{\alpha}{2} tr\left(\mathbf{X}^T\mathbf{A}^T\mathbf{V}_2^{-1}\mathbf{A}\mathbf{X}\right)\right), \tag{5}$$

where $\mathbf{Y}$ is a $d \times n$ matrix containing task fMRI time courses of d brain areas of a given subject. $\mathbf{X}$ is a $m \times n$ regressor matrix [11]. $\mathbf{A}$ is a $d \times m$ activation effect matrix to be estimated. $\mathbf{V}_1$ and $\mathbf{V}_2$ are $d \times d$ covariance matrices of $\mathbf{Y}$ and $\mathbf{A}$, respectively. $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ models the correlations between the m experimental conditions. $MN(0,\mathbf{V}_2,\alpha\mathbf{X}\mathbf{X}^{\mathrm{T}})$ denotes the matrix normal distribution, which serves as a conjugate prior of (4) with α controlling the influence of this prior on $\mathbf{A}$. We assume $\mathbf{V}_1 = I_{d\times d}$ as conventionally done. $\mathbf{V}_2$ is set to the connectivity estimates learned with our proposed approach.

3 Materials

fMRI data. Task fMRI data were collected from 60 healthy subjects at multiple imaging centers. Each subject performed 10 language, computation, and sensorimotor tasks similar to those in [12] over a period of ~5 min. RS-fMRI data of ~7 min were also collected. Data were acquired using 3T scanners from multiple manufacturers with TR = 2200 ms, TE = 30 ms, and flip angle = 75°. Standard preprocessing, including slice timing correction, motion correction, temporal detrending, and spatial normalization, was performed on the task fMRI data using the SPM8 software. Similar preprocessing was performed on the RS-fMRI data except a band-pass filter with cutoff frequencies at 0.01 to 0.1 Hz was applied. White matter and cerebrospinal fluid confounds were regressed out from the gray matter voxel time courses.

dMRI data. dMRI data were collected from the same 60 subjects. Acquisition sequence similar to [13] was used with TR = 15000 ms, TE = 104 ms, flip angle = 90°,

32 gradient directions, and b-value = 1300 s/mm^2. MedINRIA was employed for tensor estimation and fiber tractography [14]. We warped our brain parcel template (described below) onto each subject's B_0 volume to facilitate fiber count computation.

Brain Parcellation. We divided the brain into P parcels (P set to 500) to enable finer brain partitioning than facilitated by standard brain atlases (typically $P < 150$). This choice of P provides a balance between functional localization and robustness to subject variability in tractography [7]. Parcellation was performed by concatenating RS-fMRI time courses across subjects and applying Ward clustering [15]. Parcel time courses were then generated by averaging the voxel time courses within each parcel, and normalized by subtracting the mean and dividing by the standard deviation.

4 Results and Discussion

To explore the gain in learning connectivity jointly from dMRI and RS-fMRI data, we compared based on DSC the subject consistency in the support of $\hat{\mathbf{\Lambda}}_{ij}$ as estimated using our SGGM-based approach with anatomical capacity-weighted sparse penalization (SGGM-A) versus uniform sparse penalization without anatomical information (SGGM-U). To establish a baseline for comparison, we permuted 200 times the columns and rows of $\hat{\mathbf{\Lambda}}_{ij}$ learned with SGGM-A to generate a null distribution of DSC. The average DSC of both SGGM-A and SGGM-U (Fig. 1(a)) were significantly greater than that of the null distribution based on a Wilcoxon signed rank test (p-value < 0.05). Hence, the observed subject consistency was significantly above chance. Moreover, DSC of SGGM-A was significantly higher than that of SGGM-U (p-values < 0.05), thus demonstrating the benefits of multimodal integration for connectivity estimation. In addition, we compared the sensitivity in group activation detection using: 1) ordinary least square (OLS) [11] without any connectivity prior, 2) ridge regression to control overfitting [9], connectivity-informed activation model [9] with connectivity prior learned by applying 3) oracle approximating shrinkage (OAS) on RS-fMRI data only [16], 4) SGGM-U on RS-fMRI data only, and 5) SGGM-A on dMRI and RS-fMRI data jointly. We examined 21 contrasts between the 10 experimental conditions. To enforce strict control over FPR so that we can safely base our validation on increased group activation detection, max-t permutation test [17] was used. Fig. 1(b) shows the percentage of parcels detected with significant activation averaged over contrasts. In agreement with [9], adding a connectivity prior substantially improved activation detection over OLS and ridge regression. More importantly, SGGM-A was found to outperform both SGGM-U and OAS. To assess whether the enhanced detection was statistically significant, we used a permutation test. Specifically, for each permutation, we first randomly selected half of the parcels and exchanged the labels (active or non-active) assigned by SGGM-A and each of the contrasted methods in turn for a given p-value threshold. We then computed the difference in the average number of detected parcels, and performed this procedure 10,000 times to generate a null distribution. For all p-value thresholds in Fig. 2(b), the origi-

nal difference in the number of detected parcels was found to be greater than the 95th percentile of the corresponding null distribution. Hence, the detection improvement was statistically significant. Qualitatively, our approach additionally detected relevant areas adjacent to those found by considering functional information alone (Fig. 1(c)).

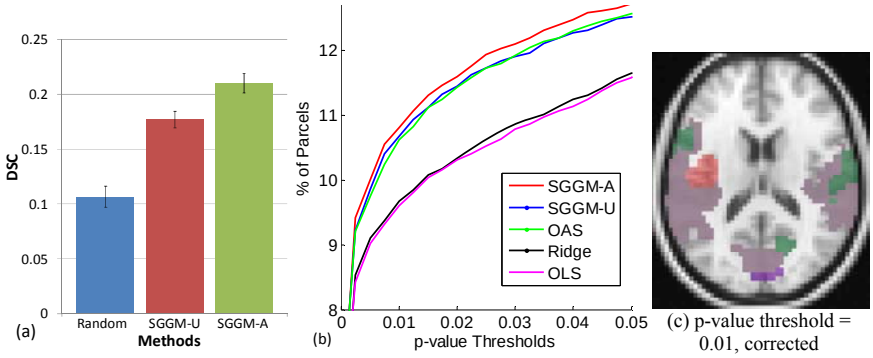

Fig. 1. Real data results. (a) Subject consistency based on DSC. (b) Percentage of parcels with significant activation averaged over contrasts vs. *p*-value thresholds. (c) Parcels detected for auditory tasks. Red = detected by SGGM-A only. Purple = detected by SGGM-A and SGGM-U. Green = detected by SGGM-A, SGGM-U, and OAS. Violet = detected by all methods.

5 Conclusions

We proposed a novel SGGM-based approach for multimodal brain connectivity inference. We showed that integrating dMRI and RS-fMRI data significantly increases subject consistency in the learned connection structure compared to analyzing RS-fMRI data alone. Enhanced group activation detection was also demonstrated. Our results thus suggest that connectivity estimated by combining anatomical and functional information may better resemble the underlying neural pathways than solely relying on functional information. A particularly important byproduct of this work is that our multimodal connectivity estimation approach in combination with our connectivity-informed activation model [9] provides a statistically-rigorous test bench for comparing different dMRI processing strategies. Quantitative evaluation, as opposed to qualitative assessment as often used in most dMRI studies, is thus facilitated.

Acknowledgments

This work was supported by the ANR grant, BrainPedia ANR-10-JCJC 1408-01, and the Berkeley INRIA Stanford grant. The data were acquired within the IMAGEN project. Jean Baptiste Poline was partly funded by the IMAGEN project, which receives funding from the E.U. Community's FP6, LSHM-CT-2007-037286. This manuscript reflects only the authors' views and the Community is not liable for any use that may be made of the information contained therein.

References

1. Fox, M.D., Greicius, M.D.: Clincial Applications of Resting State Functional Connectivity. Front. Syst. Neurosci. 4, 19 (2010)
2. Varoquaux, G., Gramfort, A., Poline, J.B., Thirion, B.: Brain Covariance Selection: Better Individual Functional Connectivity Models Using Population Prior. In: Advances in Neural Information Processing Systems. 23, 2334—2342 (2010)
3. Honey, C.J., Thivierge, J.P., Sporns, O.: Can Structure Predict Function in the Human Brain. NeuroImage 52, 766—776 (2010)
4. Damoiseaux, J.S., Greicius, M.D.: Greater than the Sum of its Parts: A Review of Studies Combining Structural Connectivity and Resting-state Functional Connectivity. Brain Struct. Funct. 213, 525—533 (2009)
5. Zhu, D., Li, K., Faraco, CC., Deng, F., Zhang, D., Guo, L., Miller, L.S., Liu, T.: Optimization of Functional Brain ROIs via Maximization of Consistency of Structural Connectivity Profiles. NeuroImage 59, 1382—1393 (2012)
6. Si, J., Pearlson, G., Caprihan, A, Adali, T., Kiehl, K.A., Liu, J., Yamamoto, J., Calhoun., V.D.: Discriminating Schizophrenia and Bipolar Disorder by Fusing fMRI and DTI in a Multimodal CCA + joint ICA Model. NeuroImage 57, 839—855 (2011)
7. Venkataraman, A., Rathi, Y., Kubicki, M., Westin, C.F., Golland, P.: Joint Modeling of Anatomical and Functional Connectivity for Population Studies. IEEE Trans. Med. Imaging 31, 164—182 (2012)
8. Hsieh, C.J., Sustik, M.A., Dhillon, I.S., Ravikumar, P.: Sparse Invers Covariance Matrix Estimation Using Quadratic Approximation. In: Advances in Neural Information Processing Systems. 24, 2330—2338 (2011)
9. Ng, B., Abugharbieh, R., Varoquaux, G., Poline, J.B., Thirion, B.: Connectivity-informed fMRI Activation Detection. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011. LNCS, vol. 6892, pp. 285—292, Springer, Heidelberg (2011)
10. Duchi, J., Gould, S., Koller, D.: Projected Subgradient Methods for Learning Sparse Gaussions. In: Int. Conf. Uncertainty in Artificial Intelligence (2008)
11. Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D., Frackowiak, R.S.J.: Statistical Parametric Maps in Functional Imaging: A General Linear Approach. Hum. Brain Mapp. 2, 189—210 (1995)
12. Pinel, P., Thirion, B, Meriaux, S., Jober, A., Serres, J., Le Bihan, D., Poline, J.B., Dehaene, S.: Fast Reproducible Identification and Large-scale Databasing of Individual Functional Cognitive Networks. BioMed. Central Neurosci. 8, 91 (2007)
13. Jones, D.K., Williams, S.C.R., Gasston, D., Horsfield, M.A., Simmons, A., Howard, R.: Isotropic Resolution Diffusion Tensor Imaging with Whole Brain Acquisition in a Clinically Acceptable Time. Human Brain Mapping 15, 216—230 (2002)
14. Toussaint, N., Souplet, J.C., Fillard, P.: MedINRIA: Medical Image Navigation and Research Tool by INRIA. In: MICCAI Workshop on Interaction in Medical Image Analysis and Visualization, pp. 1—8 (2007)
15. Michel, V., Gramfort, A., Varoquaux, G., Eger, E., Keribin, C., Thirion, B.: A Supervised Clustering Approach for fMRI-based Inference of Brain States. Patt. Recog. 45, 2041—2049 (2012)
16. Chen, Y., Wiesel, A., Eldar, Y.C., Hero, A.O.: Shrinkage Algorithms for MMSE Covariance Estimation. IEEE Trans. Sig. Proc. 58, 5016—5029 (2010)
17. Nichols, T., Hayasaka, S.: Controlling the Familywise Error Rate in Functional Neuroimaging: a Comparative Review. Stat. Methods Med. Research 12, 419—446 (2003)